Language identifying codes: remaining issues, future prospects

The work of organisations such as PARADISEC is crucially dependent on accurate and reliable identification of the languages which are represented in resources. For efficient discovery of resources to be possible, an identifying system which is accurate and stable in itself is necessary, as is wide agreement to use the system across the relevant communities (archivists and researchers from various disciplines). ISO 639-3 is such a system and acceptance of it is now widespread; this should not, however, be taken as meaning that no problems remain and in this paper we draw attention to some of the remaining issues and the potential role of Australian researchers in working towards their solution.

ISO 639-3 reflects the reality of language differentiation more or less accurately depending on the region in question. A process for requesting revisions to the codes exists and is being used quite extensively by scholars working on Australian languages. The experience thus being accumulated will be of value in future work on language identification. This process also draws attention to another area where improvement can be made: currently, the different parts of ISO 639 (639-1, 639-2 etc.) have different registration authorities. Bringing all parts of the standard together under a single registration authority would have benefits for ongoing revisions and for transparency and is therefore an important goal. Another important goal is to ensure that linguists are able to provide input to three parts of ISO 639 currently being developed:

- ISO 639-5 a proposed set of codes for identifying groupings above the level of the single language,
- ISO 639-6 a proposed set of codes for identifying linguistic entities below the level of the single language,
- ISO 639-4 will provide an account of the principles on which the various codings rest.

Australia is represented in ISO by Standards Australia, and this body has observer status in relation to ISO Technical Committee 37 which is responsible for the 639 group of standards. A group of interested scholars in Australia constitute an informal reference group for these issues (ARGILaRe: http://users.monash.edu.au/~smusgrav/ARGILaRe/) and this group is establishing ways to provide expert input. These include the establishment of a mirror committee for TC37 under the ambit of Standards Australia, ongoing involvement with international projects and endeavours, and the potential formation of a Working Group within the Research Data Alliance framework. The goal of improving access to language resources should be one which unites various research communities and therefore we are optimistic that such endeavours can and will produce valuable results.